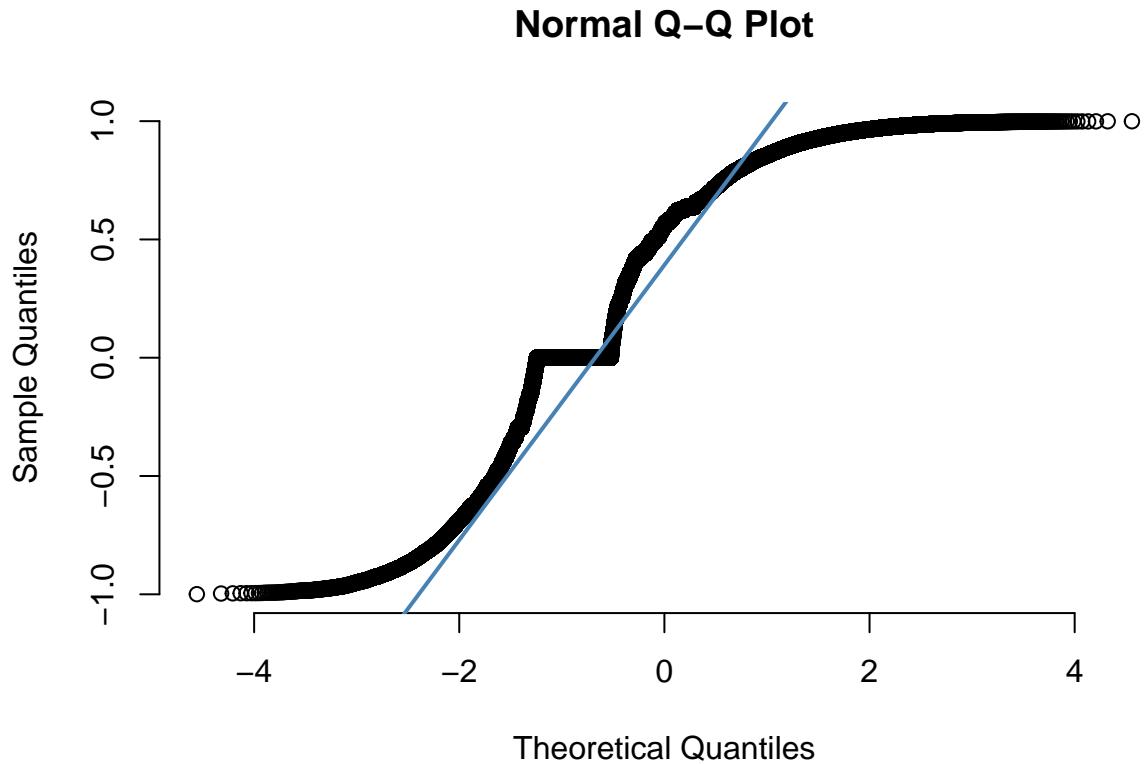


stat_analysis

2024-09-29

```
data <- read_csv('C:\\\\Users\\\\nerfm\\\\Documents\\\\Python Scripts\\\\CDC\\\\FINAL_sentiment_merged.csv')

## New names:
## Rows: 194013 Columns: 15
## -- Column specification
## ----- Delimiter: "," chr
## (4): language, source, text, location dbl (10): ...1, polarity, rating,
## wordsCount, sentiment, temperature_2m_max... date (1): time
## i Use 'spec()' to retrieve the full column specification for this data. i
## Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## * ' -> '...1'
```



```
kruskal.test(sentiment ~ factor(location), data = data)
```

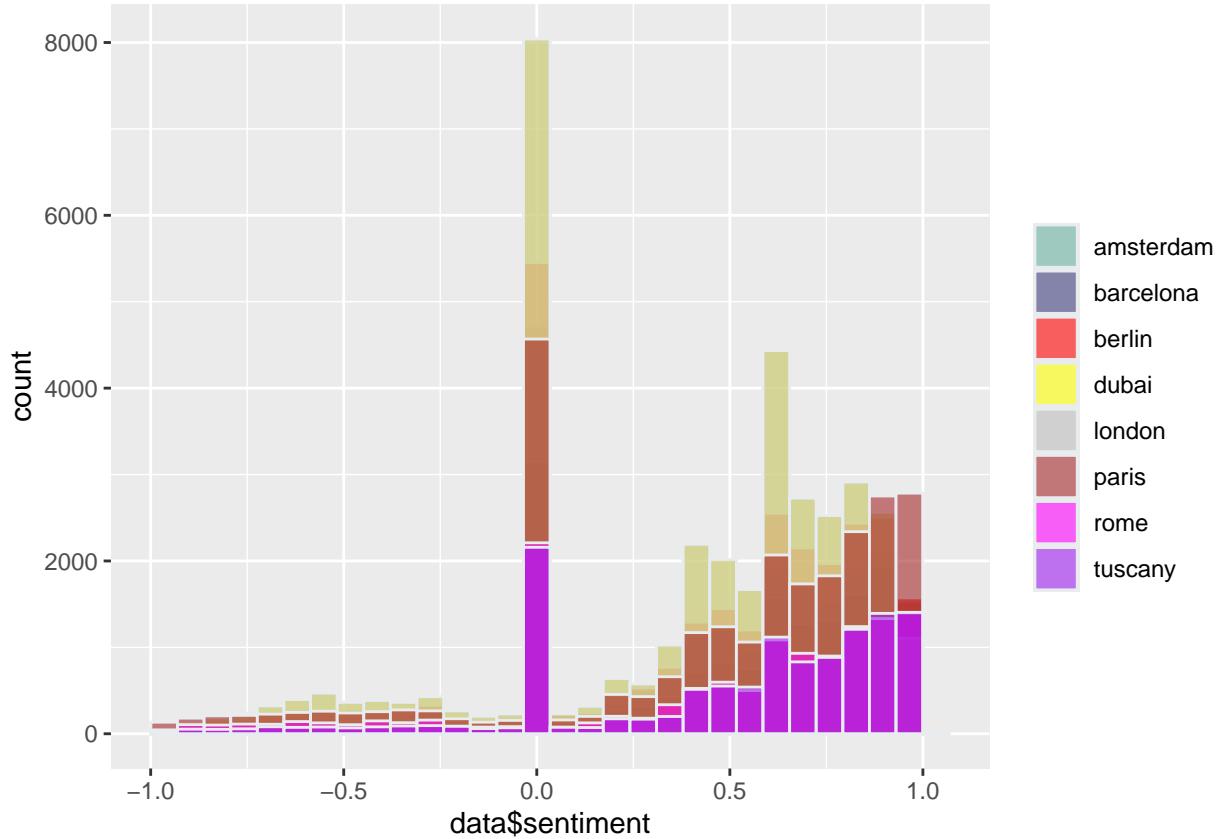
```
##
```

```
## Kruskal-Wallis rank sum test
##
## data: sentiment by factor(location)
## Kruskal-Wallis chi-squared = 2084.3, df = 7, p-value < 2.2e-16
```

```
p <- data %>%
  ggplot( aes(x=data$sentiment, fill=factor(data$location))) +
  geom_histogram( color="#e9ecef", alpha=0.6, position = 'identity') +
  scale_fill_manual(values=c("#69b3a2", "#404080",'red','yellow','gray','brown','magenta','purple')) +
  labs(fill="")
```

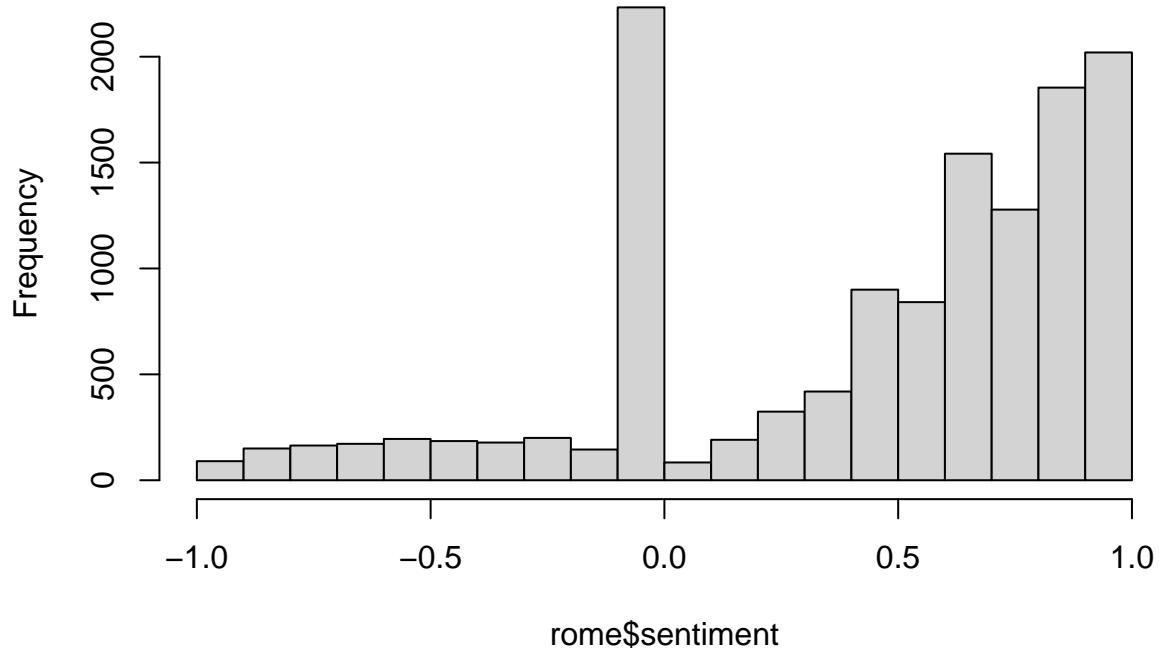
p

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.



```
rome <- data %>% filter(location == 'rome')
hist(rome$sentiment)
```

Histogram of rome\$sentiment

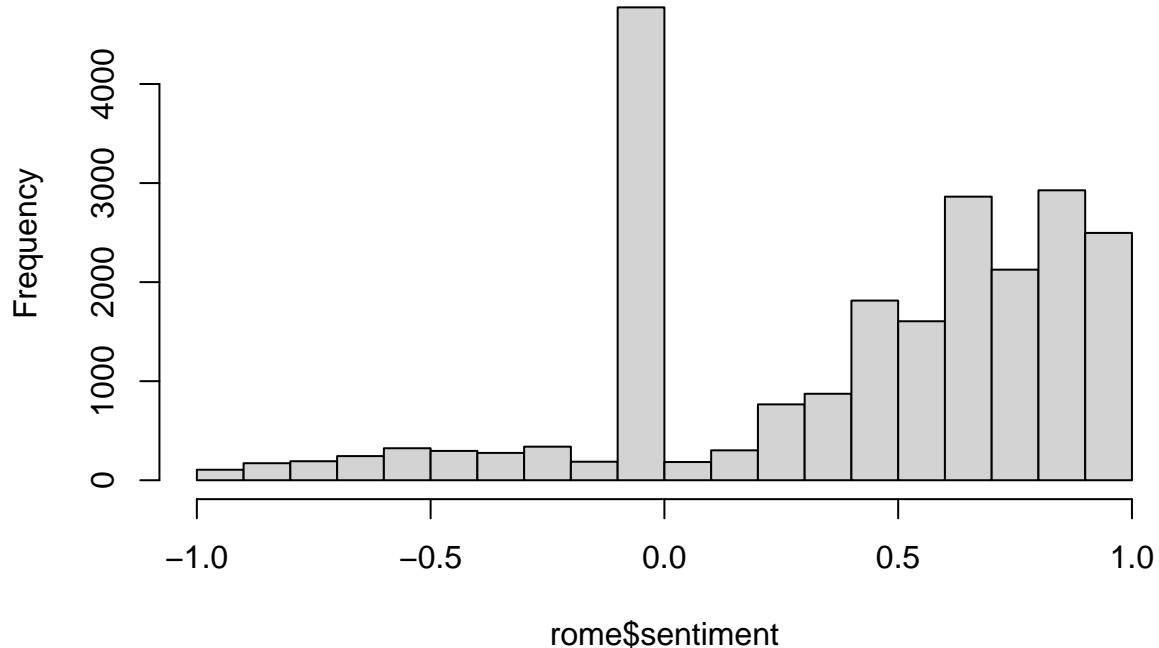


```
mean(rome$sentiment)

## [1] 0.4414225

rome <- data %>% filter(location == 'amsterdam')
hist(rome$sentiment)
```

Histogram of rome\$sentiment

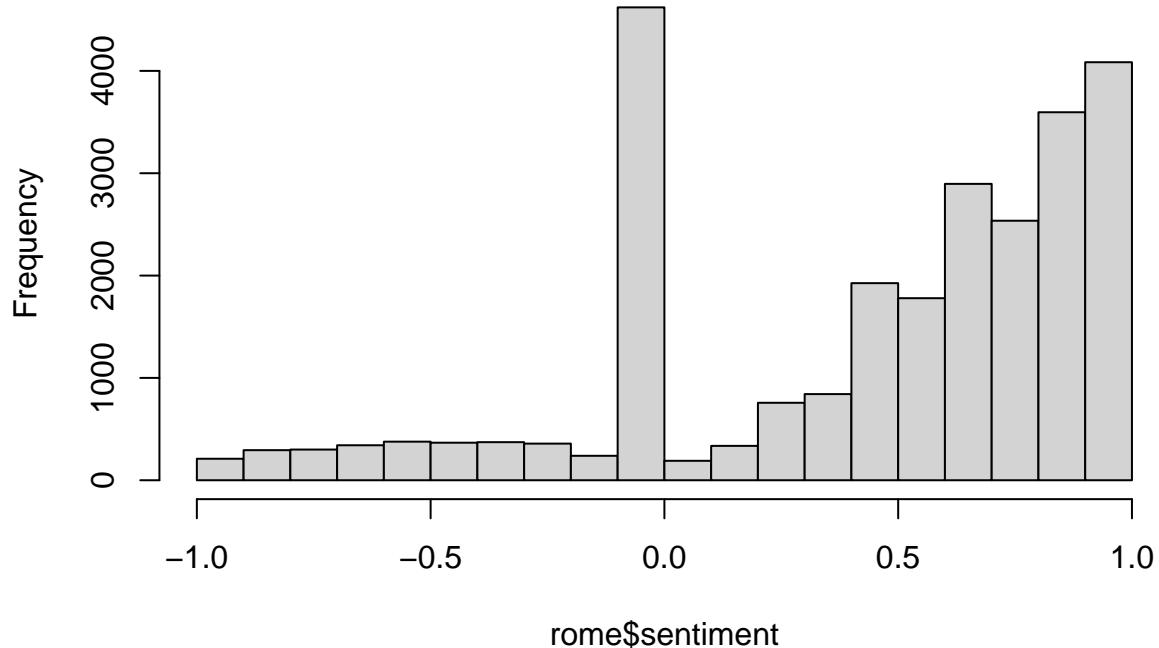


```
mean(rome$sentiment)

## [1] 0.414282

rome <- data %>% filter(location == 'paris')
hist(rome$sentiment)
```

Histogram of rome\$sentiment

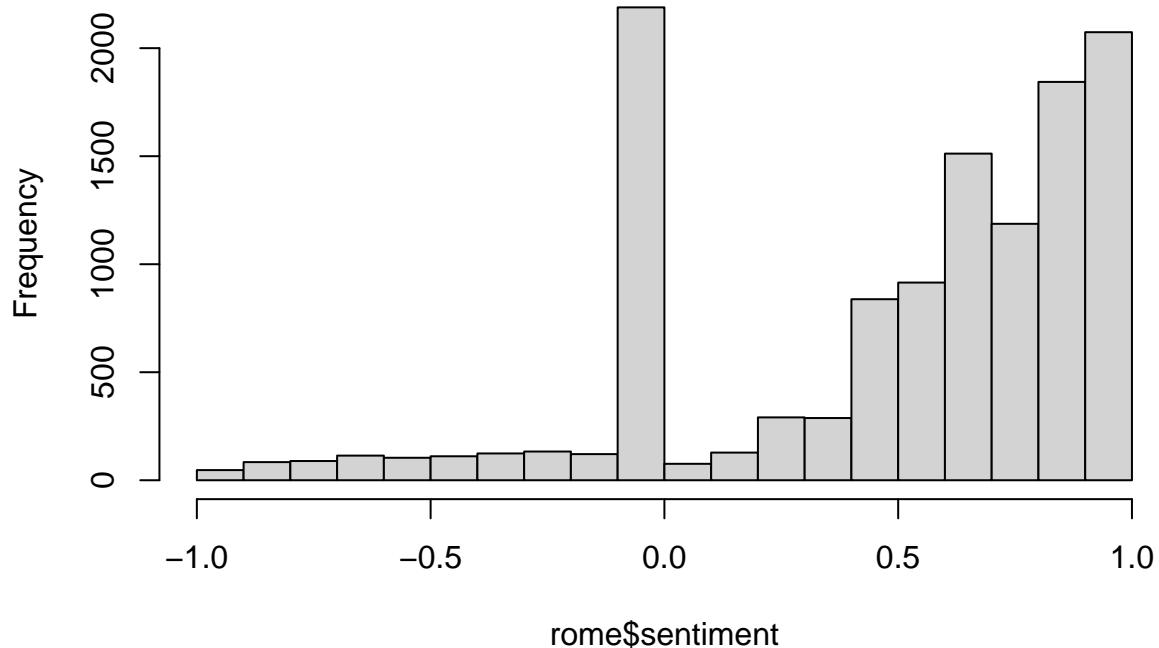


```
mean(rome$sentiment)

## [1] 0.4379912

rome <- data %>% filter(location == 'tuscany')
hist(rome$sentiment)
```

Histogram of rome\$sentiment

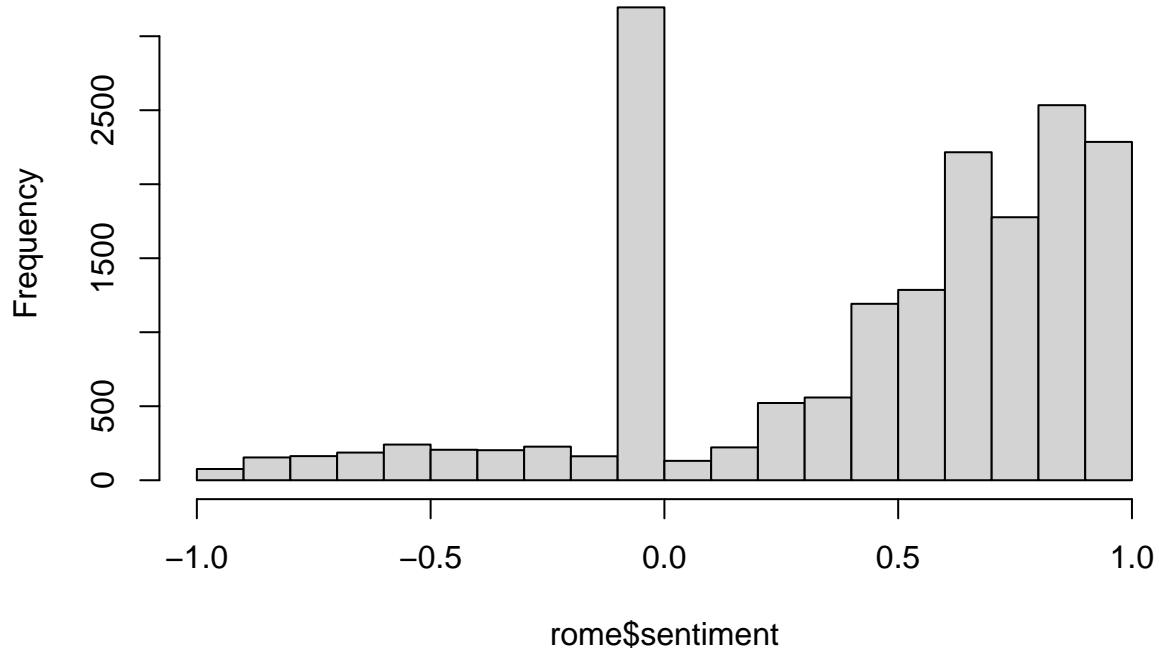


```
mean(rome$sentiment)

## [1] 0.4916195

rome <- data %>% filter(location == 'barcelona')
hist(rome$sentiment)
```

Histogram of rome\$sentiment

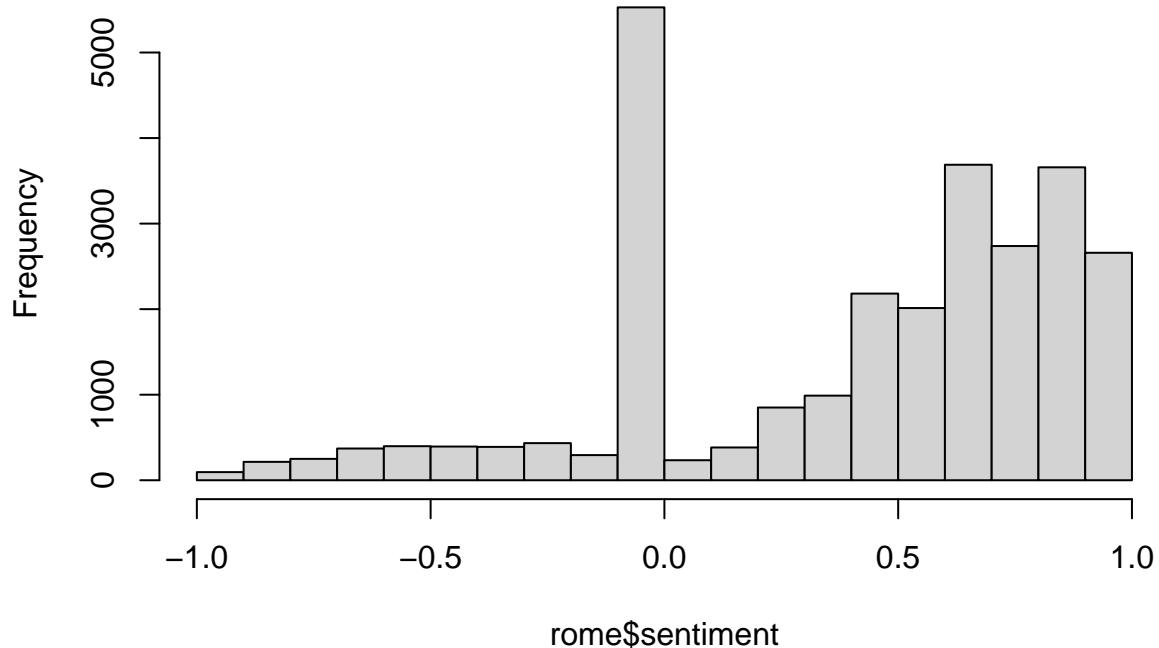


```
mean(rome$sentiment)

## [1] 0.4488232

rome <- data %>% filter(location == 'berlin')
hist(rome$sentiment)
```

Histogram of rome\$sentiment

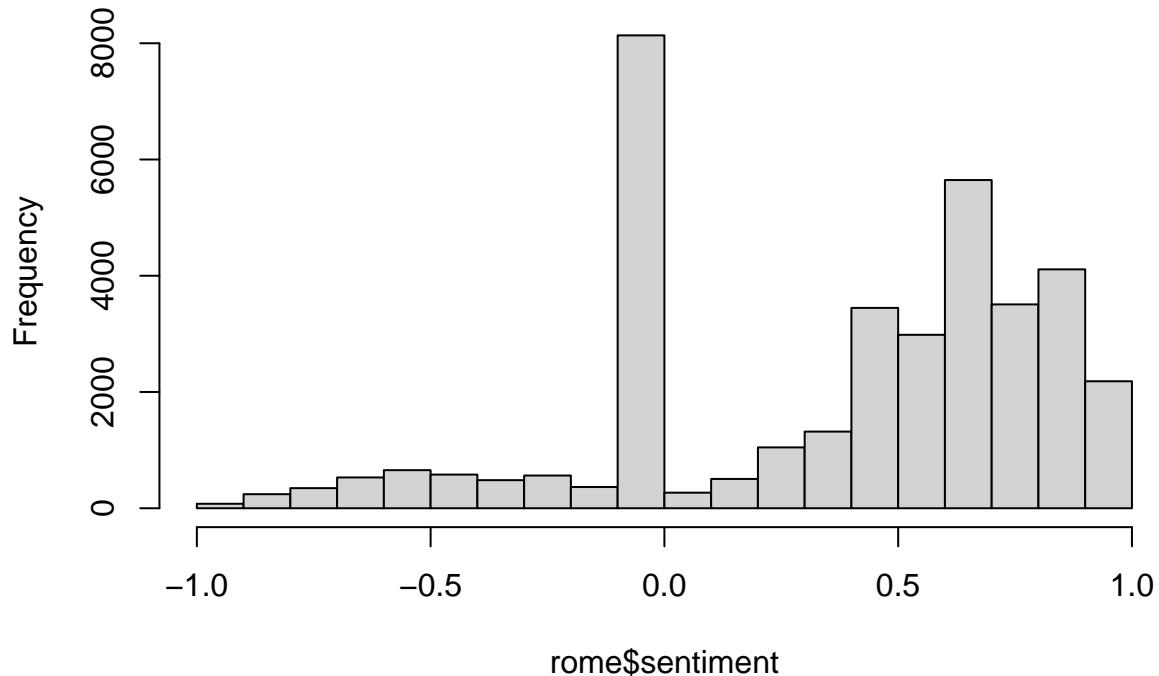


```
mean(rome$sentiment)

## [1] 0.410313

rome <- data %>% filter(location == 'dubai')
hist(rome$sentiment)
```

Histogram of rome\$sentiment

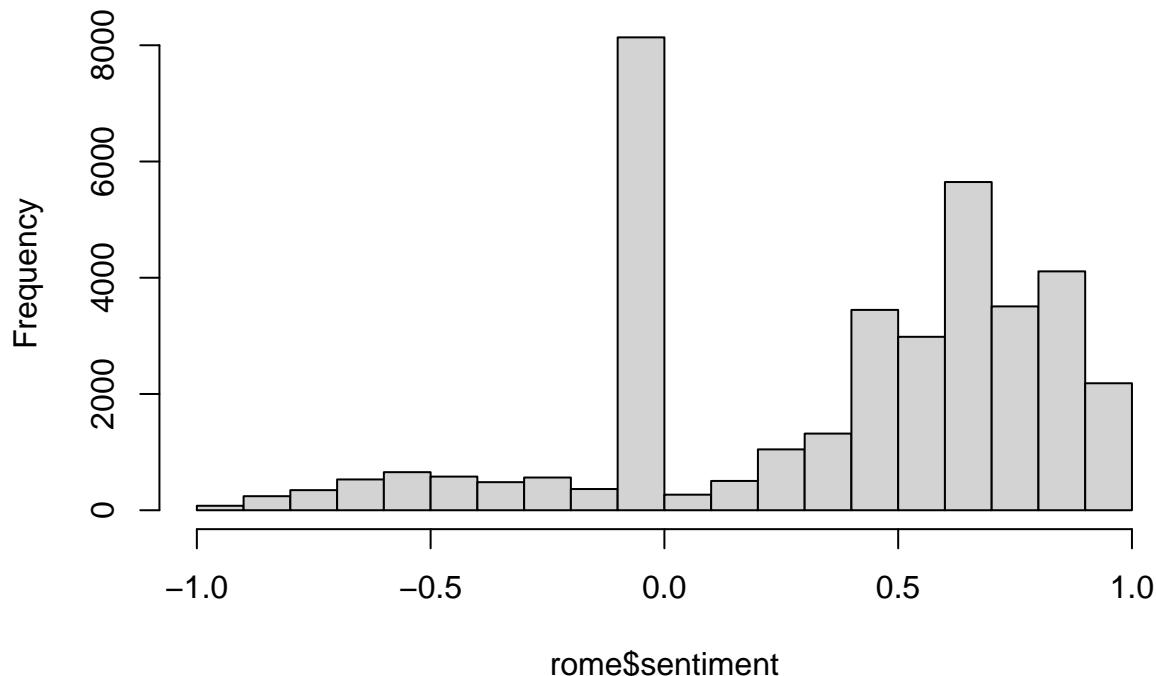


```
mean(rome$sentiment)

## [1] 0.3774755

rome <- data %>% filter(location == 'london')
hist(rome$sentiment)
```

Histogram of rome\$sentiment



```
mean(rome$sentiment)

## [1] 0.3774755

library(dunn.test)
dunn.sentiment <- dunn.test(data$sentiment,factor(data$location),kw = T, method = 'bonferroni')

## Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 2084.2615, df = 7, p-value = 0
##
##
## Comparison of x by group
## (Bonferroni)
## Col Mean-
## Row Mean | amsterda barcelon berlin dubai london paris
## -----+-----
## barcelon | -9.808150
## | 0.0000*
## |
## berlin | 1.176434 11.29589
## | 1.0000 0.0000*
## |
## dubai | 13.70211 23.31139 13.19302
```

```

##          | 0.0000* 0.0000* 0.0000*
##          |
## london | 13.70211 23.31139 13.19302 0.000000
##          | 0.0000* 0.0000* 0.0000* 1.0000
##          |
## paris  | -10.96194 -0.057638 -12.74254 -26.60253 -26.60253
##          | 0.0000* 1.0000 0.0000* 0.0000* 0.0000*
##          |
## rome   | -10.03209 -0.980524 -11.36470 -22.17285 -22.17285 -1.007271
##          | 0.0000* 1.0000 0.0000* 0.0000* 0.0000* 1.0000
##          |
## tuscany | -18.25459 -8.992683 -19.81326 -30.67267 -30.67267 -9.636605
##          | 0.0000* 0.0000* 0.0000* 0.0000* 0.0000* 0.0000*
## Col Mean-
## Row Mean | rome
## -----
## tuscany | -7.533388
##          | 0.0000*
##          |
## alpha = 0.05
## Reject Ho if p <= alpha/2

```

```
library(fpp3)
```

```

## Registered S3 method overwritten by 'tsibble':
##   method           from
##   as_tibble.grouped_df dplyr

```

```
## -- Attaching packages ----- fpp3 1.0.1 --
```

```

## v tsibble     1.1.5    v feasts      0.4.1
## v tsibbledata 0.4.1    v fable       0.4.0

```

```

## -- Conflicts ----- fpp3_conflicts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()

```

```

data <- data %>% mutate(date = as.Date(time,format = "%m/%d/%Y"))
data <- data %>% mutate(yearmonth = yearmonth(date))

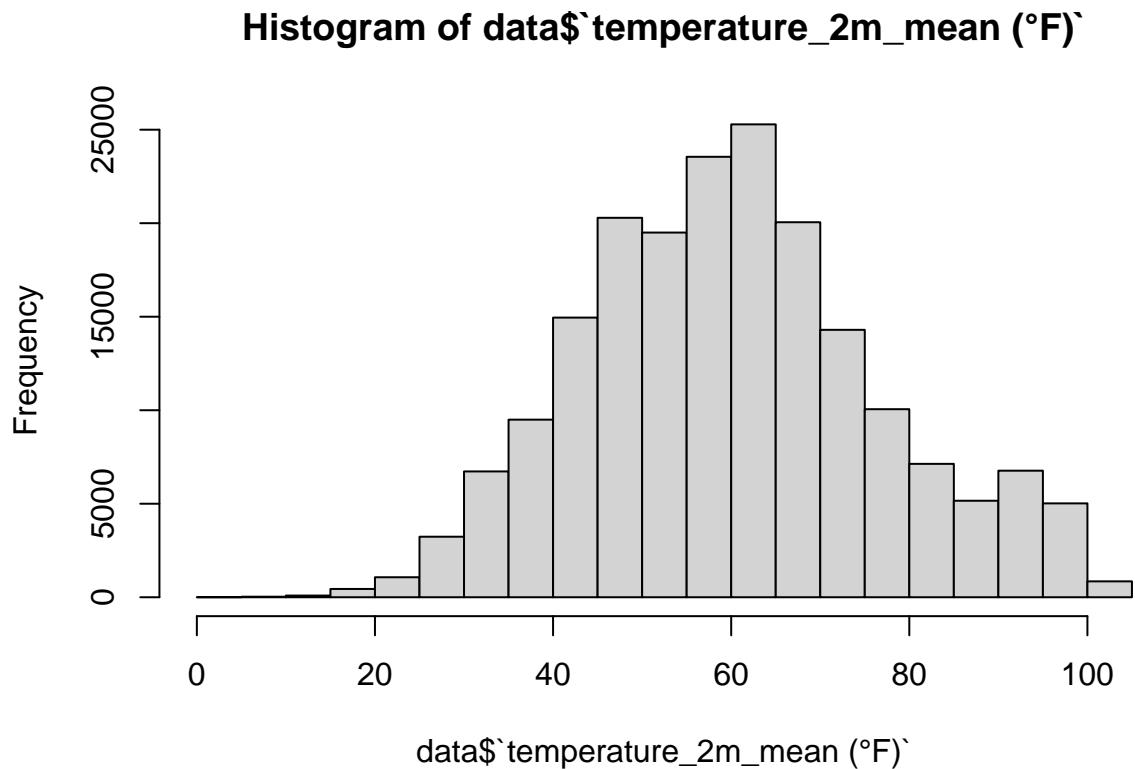
```

```

data$sentiment_cat <- cut(data$sentiment,
                           breaks=c(-1, -.1,.1, 1),
                           labels=c('Negative', 'Neutral', 'Positive'))

```

```
hist(data$`temperature_2m_mean (°F)`)
```



```

library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##   recode

## The following object is masked from 'package:purrr':
##   some

library(PMCMRplus)
leveneTest(data$`temperature_2m_mean (°F)` ~ data$sentiment_cat:factor(data$location))

## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      23  235.74 < 2.2e-16 ***
##             193989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

temp_by_sentiment <- aov(data$`temperature_2m_mean (°F)` ~ data$sentiment_cat + factor(data$location))
final <- gamesHowellTest(temp_by_sentiment)
summary(final)

## 
## Pairwise comparisons using Games-Howell test

## data: data$`temperature_2m_mean (°F)` by data$sentiment_cat by factor(data$location)

## alternative hypothesis: two.sided

## P value adjustment method: none

## H0

##          q   value   Pr(>|q|) 
## Neutral - Negative == 0 -4.142 0.0095222 ** 
## Positive - Negative == 0 -3.359 0.0461589   * 
## Positive - Neutral == 0   1.853 0.3895781 

## --- 

## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```